

Tagungsbeitrag zu:

Jahrestagung der DBG, Kommission V

Titel der Tagung:

Böden – eine endliche Ressource

Veranstalter: DBG, September 2009, Bonn

Berichte der DBG (nicht begutachtete online Publikation) <http://www.dbges.de>

Zusammenhang zwischen Rastergröße und Modellgüte für die Prognose von Bodenkarten im Maßstab 1 : 25.000

Tim Häring¹, Elke Dietz,
Christian Kölling

Zusammenfassung

Durch das Vorhandensein von flächendeckenden und hoch aufgelösten DGM stellt sich die Frage, welche Rastergröße die geeignetste für die Prognose von Bodenkarten im Maßstab 1:25 000 ist? Wir vergleichen die Modellgüte für neun Rastergrößen in fünf Regionen in Bayern und stellen fest, dass eine geringe Rastergröße nicht zu den besten Ergebnissen führt. Der Modellfehler nimmt mit zunehmender Rastergröße zunächst ab und steigt ab ca. 100m wieder an.

Die Verwendung des DGM in der Rastergröße mit dem geringsten Fehler scheint jedoch nicht ratsam, da die Heterogenität der (Boden)landschaft nicht mehr abgebildet werden kann. Ein praktikabler Kompromiss zwischen räumlicher Genauigkeit und Modellgüte ist u.E. eine Rasterauflösung von 25m.

Schlüsselwörter: digital soil mapping, Rastergröße, Modellierung, Modellgüte

1. Einleitung

Das Relief ist neben der Geologie, dem Klima und der Vegetation ein bedeutender Einflussfaktor für die Bodenbildung und Bodeneigenschaften (Seibert et al., 2007). Heute stehen Digitale Geländemodelle flächendeckend in hoher Qualität und räumlicher Auflösung zur Verfügung und werden zusammen mit daraus abgeleiteten Reliefparametern in nahezu 80% aller digitalen Bodenkartierungsprojekte für Prognosen digitaler Bodenkarten oder die Regionalisierung von Punktinformationen als Prädiktoren verwendet (Dobos and Hengl, 2009).

Die Frage, welche Rasterweite für die Prognose einer Bodenkarte im Maßstab 1:25 000 zu den besten Ergebnissen führt, ist nicht hinreichend geklärt. Untersuchungen zeigen unterschiedliche Ergebnisse. Die Wahl der geeigneten Rastergröße ist abhängig von mehreren Parametern wie dem Kartenmaßstab, der Zielgröße, den naturräumlichen Gegebenheiten und der räumlichen Ausdehnung des Untersuchungsraumes. Ziel dieser Untersuchung soll sein, die geeignete Rasterauflösung für die Prognose von Bodenkarten in unterschiedlichen Prognosegebieten in Bayern im Maßstab 1:25 000 zu finden.

2. Material und Methoden

In fünf naturräumlich unterschiedlichen Regionen Bayerns (Alpenraum, Alpenvorland, Fränkische Alb, Keuperbergland) wurden jeweils für neun unterschiedliche Rastergrößen Prognosemodelle für die Übersichtsbodenkarte ÜBK25 erstellt. Als Datengrundlage wurde das offizielle DGM der bayerischen Vermessungsverwaltung

¹ Tim Häring, Bayerische Landesanstalt für Wald und Forstwirtschaft, Sachgebiet Standort und Bodenschutz, Hans-Carl-von-Carlowitz-Platz 1, 85354 Freising, tim.haering@lwf.bayern.de

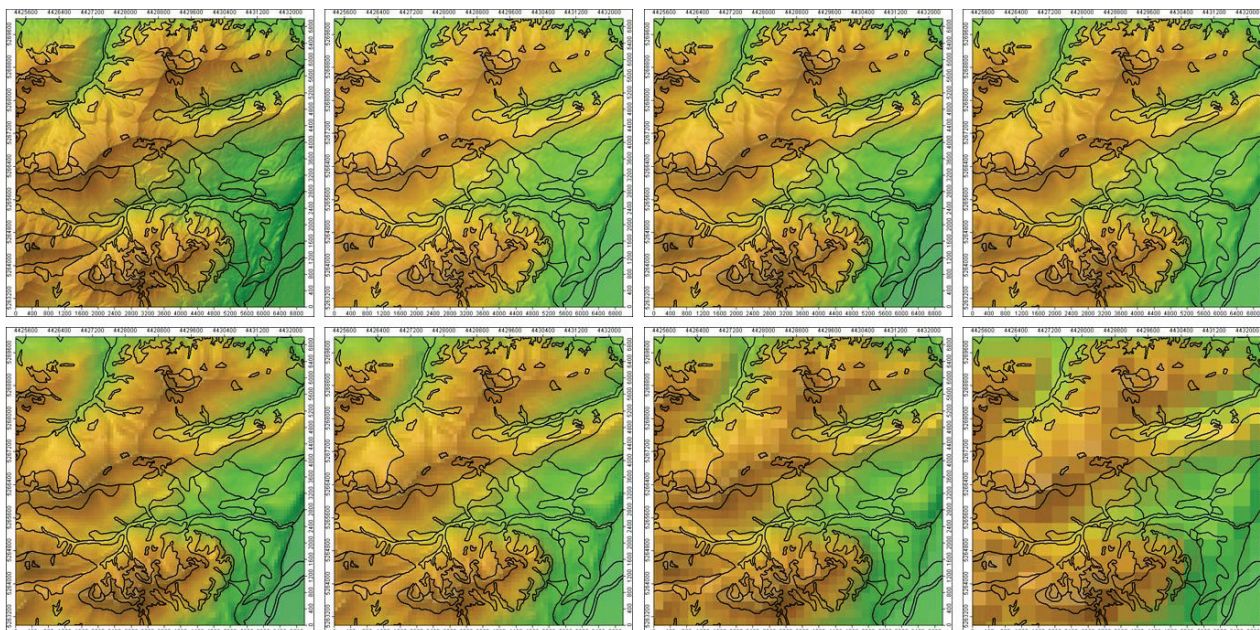


Abb. 1: Beispielhafter Ausschnitt aus dem DGM bei Garmisch-Partenkirchen für die verwendeten acht Rastergrößen (v.l.o.n.r.u.) 10m, 25m, 40m, 50m, 75m, 100m, 200m und 400m. Über dem DGM sind die Umrisse der UBK25-Einheiten zu sehen. © Datengrundlage: Bayerische Vermessungsverwaltung und Bayerisches Landesamt für Umwelt.

verwendet. Das DGM wurde durch Laserscanning-Daten erstellt, die eine Punktdichte von 1 bis 10m besitzen. Das 10m DGM wurde durch resampling auf 25m, 40m, 50m, 75m, 100m, 200m, 400m und 750m aufgelöst (Abb.1). Neben dem DGM und den daraus abgeleiteten Reliefparametern wurden klimatische und geologische Karten im Maßstab 1:25.000 und 1:200.000 als Prädiktoren verwendet.

Um die Ergebnisse sowohl innerhalb als auch zwischen den einzelnen Untersuchungsgebieten vergleichbar zu machen, wurde für jede Beispielregion das selbe Sampling und das selbe Modell verwendet. Als Sampling wurde ein *stratified random sampling* mit den einzelnen Bodeneinheiten als Strata verwendet. Innerhalb eines Untersuchungsraums wurde das Sampling beibehalten, d.h. die räumliche Verteilung und die Anzahl der Samplingpunkte ist für alle neun Rastergrößen dieselbe. Als Modell wurde *randomForest* verwendet (Breiman, 2001; Liaw and Wiener, 2002). Die

Modellvalidierung erfolgte sowohl an einem unabhängigen Testdatensatz als auch anhand des in *randomForest* implementierten, vom Trainingsdatensatz unabhängigen *out-of-bag-error* (oob). Sämtliche GIS-Analysen und die Modellierung wurden mit R (R Development Core Team, 2009) und SAGA durchgeführt.

3. Ergebnisse und Diskussion

In Abb.2 ist die Modellgüte in Form des Modellfehlers in Abhängigkeit von der Rastergröße der Reliefprädiktoren für die verschiedenen Modelldurchläufe dargestellt. Auch wenn Unterschiede zwischen den einzelnen Kurven erkennbar sind, ist ein deutlicher Trend feststellbar: Mit zunehmender Rastergröße nimmt der Modellfehler zunächst ab, wird aber in allen Untersuchungsgebieten ab einem bestimmten Wert wieder höher.

In allen Modelldurchläufen ist der Modellfehler bei einer Rastergröße von 10m relativ hoch. Zwischen 25m und

50m flachen die Kurven ab und besitzen zwischen 75m und 200m den geringsten Fehlerwert. Mit größer werdenden Rastern steigt der Fehlerwert wieder an.

Erklärt werden kann dieser Trend dadurch, dass mit zunehmender Rastergröße das Relief geglättet wird, wodurch kleinräumige Geländeformen, die in der gegebenen Auflösung nicht berücksichtigt werden, nicht mehr abgebildet werden können (Thompson et al., 2001). Letztlich verringert sich die Varianz der einzelnen Reliefattribute innerhalb einer Bodeneinheit (Abb.3). Ein Entscheidungsbaum ist deshalb in der Lage, einen Datensatz in homogenere Untergruppen zu unterteilen, was zu einem geringeren Modellfehler führt.

Überschreitet die Rastergröße einen bestimmten Wert verlieren die

Reliefprädiktoren für einen Entscheidungsbaum an Bedeutung, weil sie keine Information mehr enthalten. Mit der Geologie als alleinigen Prädiktor lassen sich die Bodeneinheiten nicht separieren. Der Modellfehler steigt wieder an.

Um die geeignete Rastergröße für die digitale Bodenkartierung zu finden, schlagen Dobos und Hengel (2009) vor, die Modellgüte mit unterschiedlichen Rastergrößen zu vergleichen und diejenige Auflösung mit dem geringsten Fehler zu verwenden. Unsere Untersuchungen zeigen, dass die Modellgüte mit zunehmender Rastergröße zunächst abnimmt. Um die räumliche Genauigkeit und die Heterogenität der Bodenlandschaft in

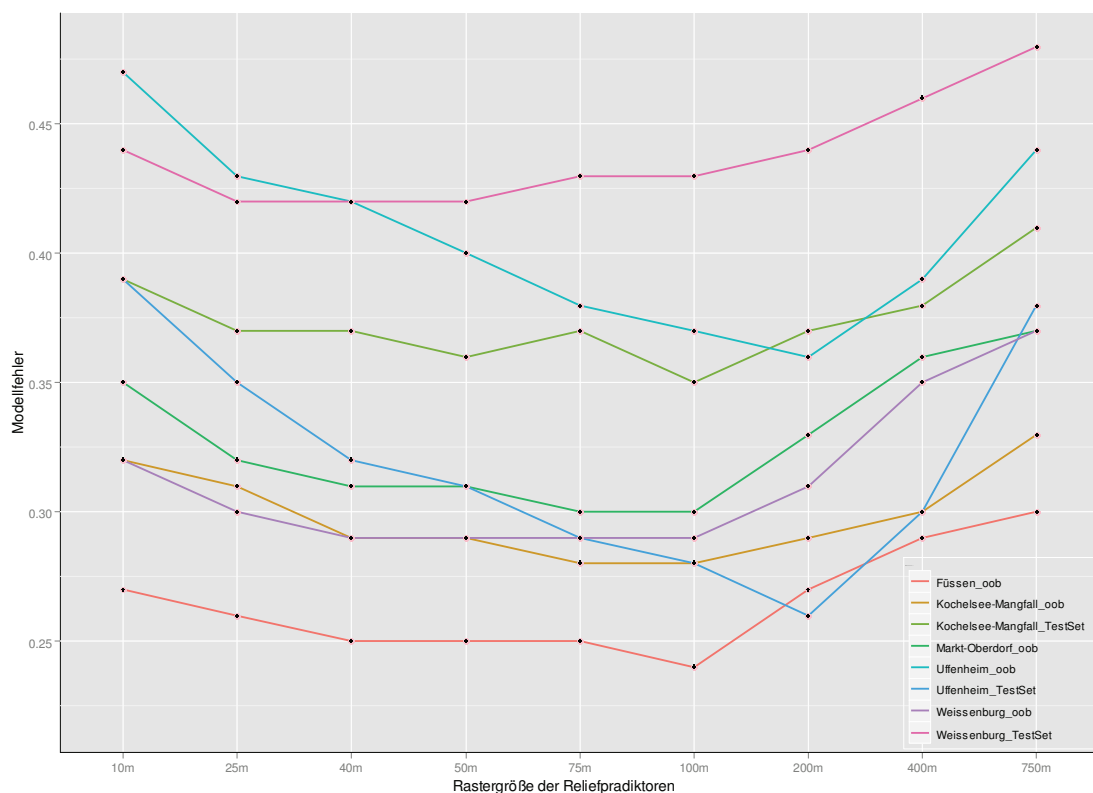


Abb. 2: Modellfehler in Abhängigkeit von der Rastergröße der Reliefprädiktoren.

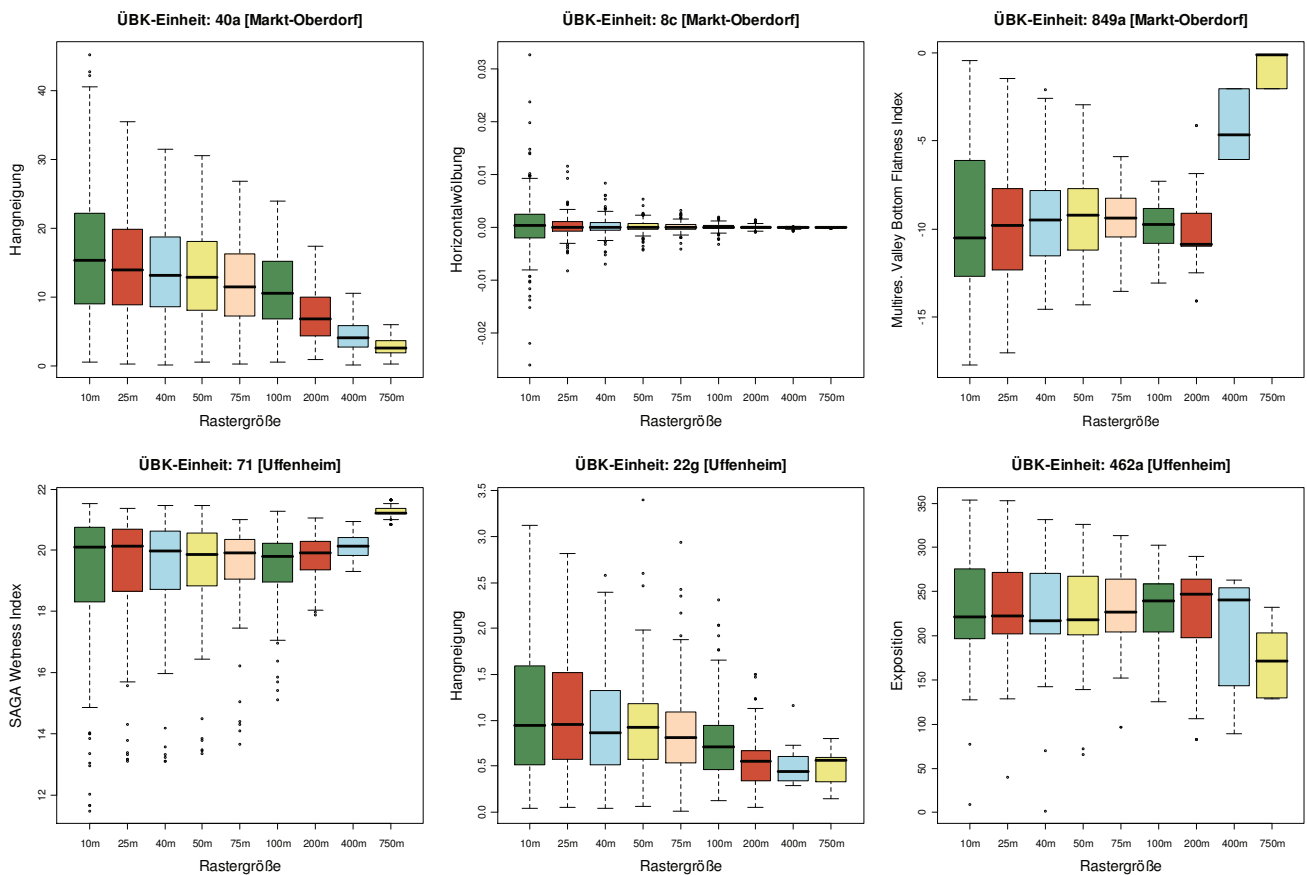


Abb. 3: Boxplotdarstellung einzelner Reliefparameter innerhalb unterschiedlicher ÜBK-Einheiten. Die Abnahme der Streuung mit zunehmender Rastergröße zeigt sich deutlich.

einem gegebenen Maßstab von 1:25.000 mit einem Modell zu greifen und letztlich in bisher unkartierte Bereiche zu prognostizieren ist es allerdings nicht vielversprechend, eine Rasterauflösung von 75m oder 100m zu verwenden. Ein praktikabler Kompromiss zwischen räumlicher Genauigkeit und Modellgüte ist u.E. eine Rasterauflösung von 25m.

4. Literatur:

Breiman, L., 2001. Random forests. Machine Learning, 45(1): 5-32.

Dobos, E. and Hengl, T., 2009. Soil Mapping Applications. In: T. Hengl and H.I. Reuter (Editors), Geomorphometry - Concepts, Software, Applications. Developments in Soil Science - Volume 33. Elsevier, Amsterdam, pp. 461-480.

Liaw, A. and Wiener, M., 2002. Classification and Regression by randomForest. R News, 2(3): 12-22.

R Development Core Team, 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.

Seibert, J., Stendahl, J. and Sorensen, R., 2007. Topographical influences on soil properties in boreal forests. Geoderma, 141: 139-148.

Thompson, J., A., Bell, J.C. and Butler, C., 2001. Digital elevation model resolution: effects on terrain attribute calculation and quantitative soil-landscape modeling. Geoderma, 100: 67-89.